JRC TECHNICAL REPORTS

Detection of abnormal behavior in trade data using Wavelets, Kalman Filter and Forward Search

Christophe Damerval

2012

Report EUR 25491 EN

European Commission
Joint Research Centre
Institute for the Protection and Security of the Citizen

Contact information
Christophe Damerval
Address: Joint Research Centre, Via Enrico Fermi 2749, 21027 Ispra (VA), Italy
E-mail: christophe.damerval@jrc.ec.europa.eu or spyros.arsenis@jrc.ec.europa.eu
Fax: +39 0332 78 5154

http://theseus.jrc.ec.europa.eu
http://ipsc.jrc.ec.europa.eu/
http://www.jrc.ec.europa.eu/

JRC72839

EUR 25491 EN

ISBN 978-92-79-26265-4

ISSN 1831-9424

doi:10.2788/46203

Luxembourg: Publications Office of the European Union, 2012

Printed in Italy

Detection of abnormal behavior in trade data using Wavelets, Kalman Filter and Forward Search

Christophe Damerval

Abstract

In this paper we address the issue of the automatic detection of abnormal behavior in time series extracted from international trade data. We motivate, review and use three specific methods, based on solid frameworks: Wavelets, Kalman Filter and Forward Search. These methods have been successfully applied to an important EU policy issue: the analysis of trade data for antifraud and antimoney-laundering, fields in which specialists are often confronted with massive datasets. Our contribution consists in an in-depth study of these approaches to assess their performance, qualitatively and quantitatively. On the one hand, we present these three approaches, underline their specific aspects and detail the used algorithms. On the other hand, we put forward a rigorous assessment methodology. We use this methodology to evaluate each method and also to compare them, on simulated time series and also on real datasets. Results show each method has its specific advantages. Their joint use could be of a high operational impact for our applications, to deal with the variety of patterns occurring in trade data.

Keywords: Time series analysis, Anomaly detection, EU trade data, Wavelets, Kalman Filter, Forward Search

Outline

We first introduce the challenge of fraud detection, data available and existing approaches. We also present methods and algorithms to detect signals of abnormal behavior in time series. Then we perform a thorough comparison of these methods, addressing both qualitative and quantitative point of view. In particular we assess them using simulated data, and also on a considerable dataset reporting real trade.

1 Introduction

1.1 Context of antifraud and fight against money laundering

The identification of fraud or money-laundering is a difficult task for which various approaches were put forward (Bolton and Hand, 2002; Fawcett, 1997). From an operational point of view, experts in the field carry out in-depth verifications (of customs declarations for instance), taking into account legislative frameworks (national, European) and specific rules. Given the vast amount of data to be processed, the adoption of automatic monitoring tools is crucial for the identification of abnormal behavior. Such tools should provide concise information on large datasets, allowing experts in the antifraud field to focus on a limited number of specific cases. The Joint Research Center of the European Commission contributes to this important EU policy issue (Fogelman-Soulie et al, 2008) using relevant sources of information, in particular a database (COMEXT, Eurostat) reporting trade flows (imports and exports of goods). The analysis of trade flows is essential for many economic issues, and of prime interest for antifraud analysts. Since a part of the EU budget comes from tax imposed on external trade, unexpected patterns in trade flows between EU and third countries can be indicators of fraud or other irregular behavior. Here we focus on this application: the identification of irregular behavior in trade data.

1.2 Trade data

The COMEXT database designed by Eurostat contains detailed information on trade between EU member states and third countries, collected by EU customs. This repository reports a huge number of records, reporting monthly-aggregated data based on customs declarations (sum of transactions over one month).

These records give the value and the volume of products exchanged between two countries, for combinations (Product,Origin,Destination). Each triplet corresponds to data

$\{x_i, y_i \in \mathbb{R}, 1 \leq i \leq N\}, i$: month, x_i: volume (in tons), y_i: value (in euros)

Let us mention dimensions of such a dataset: 27 EU countries, 291 third countries (non EU countries) and almost 15,000 products in the nomenclature, thus leading to a maximum of 115 million combinations approximately. Each record in COMEXT corresponds to one POD combination and one period (one month for monthly aggregated data). In terms of time periods, one can retrieve in COMEXT data ranging from 1988 to present, with differences depending on countries: since 1988 for the EU15 Member States, since 2004 for the 12 New Member States. The integration of new Member States lead to a sharp increase: around 20 million records per year until 2003 and more than 40 million from 2004 onwards. In practice such a dataset does not report all possible POD combinations at each time point: missing values are frequent. This can be explained by the absence of trade, too low quantities or unavailable data. Let us note such data can be considered using either a national approach (considering each EU Member State individually) or a EU-oriented approach. In the first case, data report trade between one EU country and a third country (e.g. imports from US to FR). In the second case, all EU countries are treated as one entity: this results in trade between EU and third country (e.g. exports from EU to US) – country-aggregated data. We point out that trade flows significantly vary depending on the POD considered: certain products are more traded than others, while certain countries have a preeminent role due to the size of their economy. Furthermore, some trade flows can be subject to seasonality: for instance agricultural products are naturally more traded during certain seasons, which results in a 12-month seasonal evolution. Some products are subject to economic demand development, which results in long-term increases in trade quantities. Generally several types of fluctuations can be observed, since trade is subject to a variety of economic factors. So performing a robust analysis of such heterogeneous data constitutes a great challenge.

1.3 Signals of abnormal behavior – Methods for anomaly detection

Such data can be analyzed using two main approaches: either statistical regressions on scatterplots representing quantity and value, or time series analysis (evolution over time). According to antifraud specialists, sudden changes of quantity over time can be related with abnormal behavior within trade data. To identify such signals, we focus here on time series representing the evolution of the quantity over time, using monthly-aggregated and country-aggregated data. In this regard there exist in the literature many methods for anomaly detection in time series. These come from domains such as statistics (Fox, 1972; Barnett and Lewis, 1994; Abraham and Chuang, 1989), signal processing (Soule et al, 2005), machine learning (Salvador and Chan, 2005; Ma and Perkins, 2003; Geurts, 2001), data mining (Keogh et al, 2002; Caudell and Newman, 1993; Basu and Meckersheimer, 2007; Chandola et al, 2009). Ideally the applied methods should be: based on solid frameworks; able to deal with phenomenons such as non-stationarity and seasonality; able to identify abnormal behavior (singularities, outliers); motivated by the applicative context (anti-fraud, anti-money laundering).

1.4 Used methods in the field – Rationale for their comparison

To identify abnormal behavior, specialists in the fight against fraud and money laundering were confronted with two major issues: increasing volumes of data to be processed and lack of ready-to-use data processing tools. To overcome such difficulties, statistical and signal processing techniques were applied. First methods focused on particular fraud cases, for instance cases in which trade flows suddenly increased at known dates. Then generic methods were used, like detecting the major peaks of the first derivative with respect to time. Later, more advanced tools were put forward: Wavelets (WL), Kalman Filter (KF) and Forward Search (FS). These methods produced interesting results according to specialists in the field, who are looking for indicators of potential fraud. In particular, they showed an interest for such methods to the extent they produce interpretable results and are computationally efficient.

The comparison of methods is a crucial issue in many applications, both qualitatively and quantitatively. Such comparisons often raise issues for the different ingredients used. Here all three methods detect times at which observations are seen as abnormal, which are very similar in simple cases, less in more complex ones. Let us briefly outline the underlying logic concerning WL, KF and FS (this will be later detailed). WL focus on local analysis using a multiscale framework (continuous wavelet transform). Although no specific model is assumed, the signal detection assumes a Gaussian distribution of errors. KF assumes the evolution with time is governed by a linear dynamical system, with errors following a Gaussian distribution. FS assumes data follows a model with Gaussian errors possibly contaminated by outliers. It is adapted (without being limited) to time series subject to seasonality – when the seasonality period is known *a priori*. The fitted model being robust to outliers, the signal detection relies on the difference between model predicted by normal observations and abnormal observations (outliers).

Let us now higlight links between fraud detection and anomaly detection. Fraud cases cover a variety of situations, for instance: stockpilling, fraud in export refunds, evasion of import duties, deflection of trade, smuggling, trade-based money laudering. To a certain extent, these correspond to specific patterns in trade data (outliers, upward or downward spike, level-shift). Moreover complex fraudulent operations can lead to singular patterns for several entities: an upward spike for (P_1, O_1, D_1) and a downward spike for (P_1, O_2, D_1), a systematic underpricing for a group of entities. Although the methods presented here were not designed originally for specific fraud cases, they proved to be well adapted for the identification of fraudulent behaviour in trade data (detecting anomalies, ranking them with an adequate measure, possible characterization). Since each method proved to be relevant for this application, one wonders how the methods behave on specific patterns and on real data. The computational aspects in terms of cost and time are also important elements. Besides, one asks what compromises can be made, i.e., what results should be expected or missing if only one method is used. This would allow end-users to choose which method they should use according to their needs.

2 Detection of signals of abnormal behavior within time series

We present here different methods able to process time series and to extract a certain number of signals of abnormal behavior. Let us first precise this notion: considering a method M – either wavelets (WL), Kalman Filter (KF) or Forward Search (FS) – and one time series, a signal of abnormal behavior is a time at which M detects a specific pattern (depending on the method) along with additional information such as a measure of strength (also depending on the used method). The data we consider here is made of time series

$$(t_i, x_i), t_i \in \mathbb{N} \quad x_i \in \mathbb{R} \quad i = 1..N \tag{1}$$

Each time series is processed independently from the others. We present the general framework of each method, emphasizing its relevance. We describe the algorithm used to extract signals from time series. We also give illustrations and comments on the detected patterns.

2.1 Wavelets (WL)

Wavelets are a powerful tool for data processing. This comes with mathematical properties as well as efficient algorithms. In particular a useful tool for multiscale analysis is the Continuous Wavelet Transform (CWT), defined as

$$\forall u \in \mathbb{R}, \forall s > 0 \quad Wf(u,s) = \frac{1}{\sqrt{s}} \int_{\mathbb{R}} f(t)\psi\left(\frac{t-u}{s}\right) dt \tag{2}$$

with $f : \mathbb{R} \to \mathbb{R}$ (analyzed function) and $\psi : \mathbb{R} \to \mathbb{R}$ a wavelet function (analyzing function) – see Mallat (1989); Mallat and Hwang (1992) for details. Starting from data (t_i, x_i), algorithms based on either filterbanks or fast Fourier transforms provide efficient computations of the CWT at any chosen scale $s > 0$, see Beylkin et al (1991); Strang and Nguyen (1996). Let us denote $W(t_i, s)$ these computed values. The CWT makes up a representation of the data, using a position $u \in \mathbb{R}$ and a scale $s > 0$. A well-known application of the CWT is the detection of singularities, defined as locations at which the response $|u \mapsto W(u,s)|$ attains a local maximum and surpasses a certain threshold.

This leads to the set of singularities defined as

$$\mathcal{S}_{WL} = \left\{ t_k, |W(.,s)| \text{ local max. at } t_k \text{ and } |W(t_k,s)| > \frac{\text{MAD}}{0.6745} \cdot T_{WL} \right\} \quad (3)$$

where MAD is the median absolute deviation of the wavelet coefficients, and T_{WL} a parameter to be chosen (typical value $T_{WL} = 5$). The use of the quantity MAD/0.6745 allows to surpass the noise level (Donoho and Johnstone, 1994). See also Hoaglin et al (1983) on the use of the median absolute deviation in robust statistics. Another application of wavelets is the computation of pointwise Lipschitz regularity (Jaffard and Meyer, 1996; Benassi et al, 1998). Let us recall this value of regularity is the exponent $\alpha \in \mathbb{R}$ appearing in the expression

$$|f(t) - f(t_0)| \leq C|t - t_0|^{\alpha} \quad (4)$$

where $f : \mathbb{R} \to \mathbb{R}$, $t_0 \in \mathbb{R}$ and $C > 0$, for all t in a neighborhood of t_0. This value $\alpha \in \mathbb{R}$ should not be confused with the significance level of a statistical test. Numerically this value of regularity can be computed with wavelets, by performing a linear regression at fine scales using the formula

$$\log W(u,s) = \alpha \log s + D \quad (D \in \mathbb{R}) \quad (5)$$

In practice the estimated value of regularity allows to quantify how regular or singular a pattern is: this value indicates the sharpness of a spike. This comes from the fact the regularity α is a robust characteristic value (Andersson, 1997; Damerval and Meignen, 2009). A positive value denotes a regular pattern (like a smooth evolution), a value close to zero a level-shift (like a Heaviside step) and a value close to -1 a spike (like a Dirac impulse). In practice this value of regularity is generally comprised between -2 and 2. For illustration purposes, we represent on Figure 1 a time series representing the evolution of a quantity over time, the Sombrero wavelet $\psi(t) = (1 - t^2) \exp\left(t^2/2\right)$ (used in all our experiments), and the CWT $W(u,s)$ defined in equation (2). We also represent the response (modulus of the CWT) with respect to time, and the regularity seen in equation (4) at each time instant.

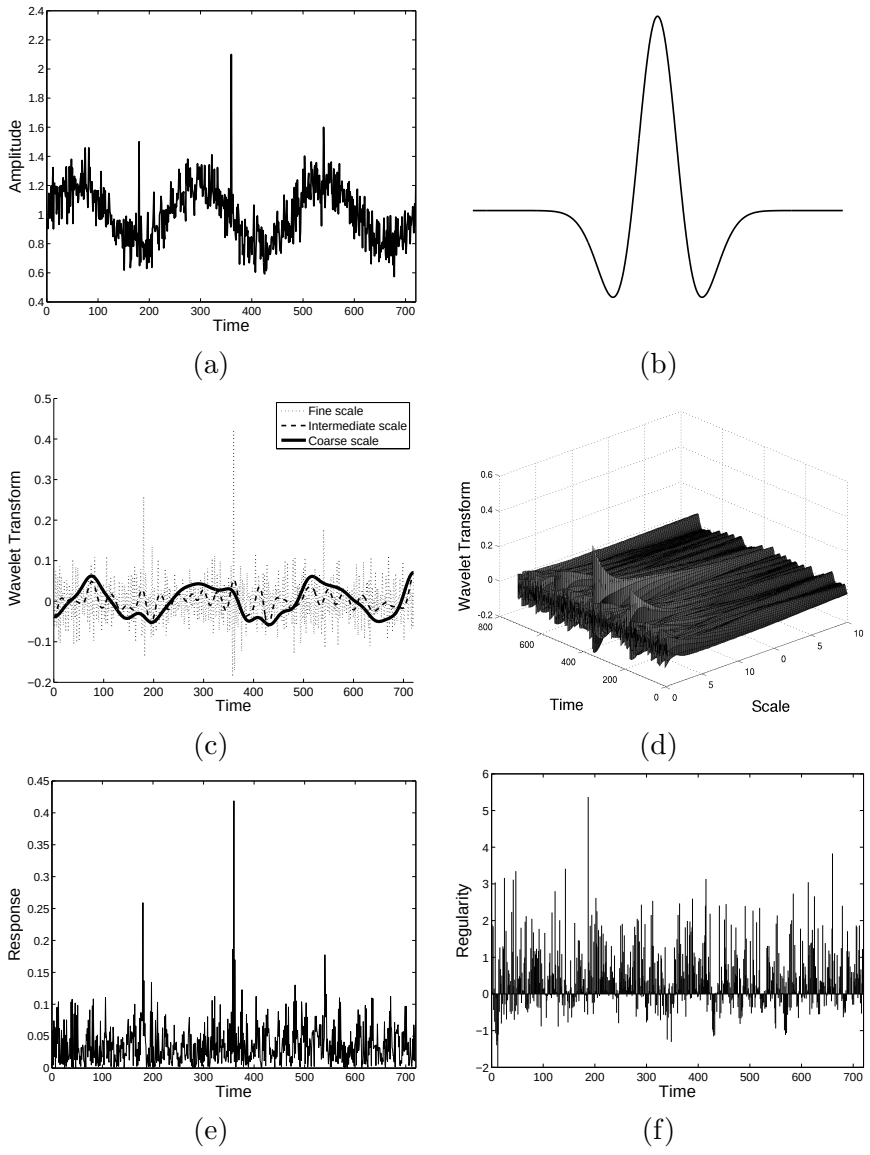


Figure 1: Illustration of wavelet approach: (a) Data: time series representing the evolution of a quantity over time; (b) Sombrero wavelet (second derivative of Gaussian), used in all our experiments; (c) Wavelet transform $u \mapsto W(u, s)$ at different scales s (fine, intermediate, coarse); (d) Wavelet transform $(u, s) \mapsto W(u, s)$ in surface representation (3D view); (e) Wavelet response $u \mapsto |W(u, s = 1)|$; (f) Values of regularity α estimated at each time.

Algorithm. Here the signals we will identify are times at which the WL method detects a singularity in the evolution of a time series. We also extract additional information (measure of strength and value of local regularity)

1. Compute the continuous wavelet transform using the second derivative of Gaussian as wavelet and three fine scales $s_1 = 1, s_2 = 2, s_3 = 3$:

$$\text{Data } (t_i, x_i)_{i=1..N} \longmapsto \text{Wavelet transform } W(t_i, s_j)_{i=1..N, j=1..3} \quad (6)$$

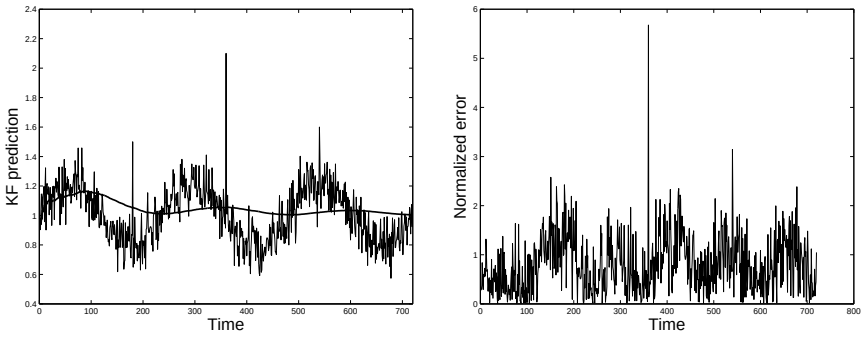
2. Identify the set of singularities $\mathcal{S}_{WL}$, using the scale $s_1 = 1$ in equation (3)

3. Compute values of regularity at each singularity location $u \in \mathcal{S}$, performing a linear regression on formula (5) using three scales $s_1 = 1$, $s_2 = 2$, $s_3 = 3$

4. Extract the following features

$$\left\{ \begin{array}{lcl} \text{Singularity location} & : & t_k \quad (k \in 1..N) \\ \text{Response} & : & |W(t_k, s_1)| \\ \text{Value of regularity} & : & \alpha_k \in \mathbb{R} \\ \text{Type of pattern} & : & \text{"spike", "level-shift" or "regular pattern"} \\ & & \left\{ \begin{array}{ll} \text{spike} & \text{if } \alpha_k < -1/2 \\ \text{level shift} & \text{if } \alpha_k \in [-1/2, 1/2] \\ \text{regular pattern} & \text{if } \alpha_k > 1/2 \end{array} \right. \end{array} \right. \quad (7)$$

The rule to determine the pattern type arises from the characteristic aspect of the regularity α (see above-mentioned theoretical values). Besides we use a measure of strength that is normalized so as to be scale invariant: two time series containing the same pattern should lead to the same strength, even if they have different amplitudes. Besides the choice of used wavelet has some importance. For analysis purposes with the CWT (case here), a classical choice is the n-th derivative of Gaussian (with n positive integer). Such a choice allows to identify (with the presented approach) the peaks of a n-th derivative of the time series. Here we used the second derivative of Gaussian (Sombrero wavelet): as a symmetric and infinitely derivable function, it allows to detect precisely singularities and provide a good regularity estimation. So the wavelet approach can be seen as a generalization of the basic approach consisting in detecting the peaks of the first derivative.



(a) (b)

Figure 2: Illustration of Kalman Filter approach: (a) Data (time series representing the evolution of a quantity over time) and prediction obtained thanks to Kalman Filter approach; (b) Normalized error (at each time) between data and KF prediction.

2.2 Kalman filter (KF)

The Kalman Filter is a classical prediction method in time series analysis, which was applied to various industrial applications (Durbin and Koopman, 2001; Kalman, 1960; Peng and Aston, 2010). The KF focuses on the temporal aspect of time series, which allows to process and predict time series progressively. This can be useful in real-time and on-line applications. The detection of abnormal behavior will be performed comparing data with KF prediction. Let us consider the monodimensional model

$$x_i = a_i + \varepsilon_i \quad (i = 1..N) \quad \varepsilon_i : \mathcal{N}(0, \sigma_1^2) \tag{8}$$

$$a_{i+1} = a_i + \eta_i \quad (i = 1..N) \quad \eta_i : \mathcal{N}(0, \sigma_2^2) \tag{9}$$

in which data $(x_i)_{i=1..N}$ are modelled as the sum of a noise and a state $(a_i)_{i=1..N}$ following a random walk. The predicted state a_{i+1} only depends on noise and the current state a_i. All noises are assumed with zero mean and constant variance. From the computational point of view, the initial state a_0 can be obtained using optimization techniques, whereas the update from a_i to a_{i+1} can be performed explicitly. A strong point of the KF lies in its capacity to estimate the state prediction and variance that formally write as

$$x_i^p = E[a_i | x_{1..i-1}] \tag{10}$$

$$v_i = Var(a_i | x_{1..i-1}) \tag{11}$$

This allows to define the normalized prediction error

$$e_i = \frac{x_i - x_i^p}{\sqrt{v_i}} \tag{12}$$

which behaves as a Gaussian distribution with zero mean and variance equal to one. This quantity turns out as relevant for the detection of abnormal behavior: provided it surpasses a certain threshold, this indicates a sharp change in the evolution of the time series. The set of sharp changes are defined as

$$\mathcal{S}_{KF} = \{t_k(k \in 1...N), e_k > T_{KF}\} \tag{13}$$

where T_{KF} is a threshold parameter to be chosen (typical value $T_{KF} = 3$). To illustrate we represent on Figure 2 the KF prediction and the normalized error associated to the time series previously seen in Figure 1. **Algorithm**. Here the signals we identify are times at which the data and the KF prediction significantly differ. Such signals correspond to sharp changes in the evolution of the time series.

1. Compute KF prediction, residuals and normalized error.

2. Extract the following features

$$\begin{cases} \text{Sharp change location} & : & t_k \quad (k \in 1..N) \\ \text{Prediction} & : & x_k^{predict} \\ \text{Residual} & : & r_k = x_k - x_k^{predict} \\ \text{Normalized error} & : & e_k > 0 \end{cases} \tag{14}$$

Remark: the KF approach can be efficiently implemented using the State Space Models Peng and Aston (2010). In terms of perspectives we mention that more complex models (non-Gaussian, non-linear) and robust versions of KF (using techniques such as reweighting) offer perspectives that are potentially relevant for our application.

2.3 Forward Search (FS)

The Forward Search is a flexible approach in robust statistics. In particular it allows to perform very robust regressions, presenting a specific adaptiveness to the data (Atkinson et al, 2004; Atkinson and Riani, 2000; Rousseeuw and Leroy, 1987). This tool turns out as efficient for different applications such as outlier detection, model selection and clustering. Here we process time series with FS using a model integrating a seasonal trend (Riani, 2004). This approach allows to evidence outliers. We will also extract additional information for every outlier: residual and measure of outlyingness. Let us recall the main steps of the Forward Search (Atkinson et al, 2004). First choose an initial subset free of outliers: this can be done using least median of squares regression (alternatively least trimmed squares) Then add progressively observations by selecting those corresponding to the smallest squared residuals. Finally monitor the evolution of the standardized residuals with respect to subset size. This allows to identify normal units – on which a classical linear regression can be fitted – and outliers (corresponding to large standardized residuals). For illustration purposes we represent on Figure 3 the fitted model by FS and the monitoring of residuals corresponding to the time series previously used with WL and KF – see Fig.1 and 2.

A strong point of the FS lies in its ability to order the data, from units rather following the model until units more likely to be outliers. This allows to evidence complementary subsets of normal units and outliers, using a test size which is ts-simultaneous ($ts \in [0, 1]$). Let us recall the notion of test size: considering a large number of datasets free of outliers, the FS will identify outliers in a fraction of them (on average equal to ts). For instance if we choose $ts = 0.01$, on average 1% of these datasets will be identified as containing at least one outlier. Since we are interested here in the detection of abnormal behavior, it is natural to consider the set

$$\mathcal{S}_{FS} = \{t_k \text{ identified as outliers by FS using a test size } T_{FS}\} \qquad (15)$$

where the parameter T_{FS} is chosen by the user (typical value $T_{FS} = 0.01$).
Computing a measure of outlyingness. The choice of the test size ts has an impact on the number of detected outliers but not on their strength. Let us explain how to compute such a measure of strength: once outliers have been identified by FS, we carry out the following steps: first fit a linear regression using normal units, second compute deletion residuals on this outlier (with appropriate formula, see Atkinson et al (2004) for details) and third perform a statistical t-test (Student test). This t-test quantifies the agreement of the outlier with the set of normal units, giving a p-value in

$[0, 1]$ measuring the outlier strength: a value close to 0 indicates a strong outlier. Let us now describe how we extract signals with FS in one time series.

Algorithm. Here the signals we identify are times corresponding to outliers detected by the Forward Search.

1. Perform FS on one time series using the approach proposed in Riani (2004). This leads to two subsets made respectively of normal units and outliers, denoted respectively $\mathcal{N}$ and $\mathcal{O}$.

2. Using the set of normal units $\mathcal{N}$, we obtain a fit taking into account a trend, either linear or seasonal. Using the set of outliers $\mathcal{O}$, we compute for each outlier the residual and a p-value (using a t-test). These quantify the outlyingness of each outlier.

3. Extract the following features

$$\begin{cases} \text{Outlier location} & : & t_k \quad (k \in 1..N) \\ \text{Fitted value} & : & x_k^{fit} \\ \text{Residual} & : & r_k = x_k - x_k^{fit} \\ \text{p-value} & : & p_k \in [0, 1] \end{cases} \tag{16}$$

2.4 Notes on practical implementation

Each of the presented methods can be implemented efficiently in Matlab environment. Concerning WL, we mention there are toolboxes available to the scientific community (Wavelab, 1992). For our experiments we developed a specific wavelet analysis toolbox, with an emphasis on the CWT and robust regularity estimation. Concerning KF, we used the SSM toolbox (Peng and Aston, 2007), which is an efficient and general tool performing at state-of-the-art. Concerning FS, we mention the reference toolbox FSDA, Forward Search and Data Analysis (FSDA, 2011). We used this toolbox, and designed specific functions to process time series taking into account seasonality. Finally we emphasize that all computations are performed using the same software environment (Matlab) and efficient implementations performing at state-of-the-art. This ensures the validity of further comparisons.

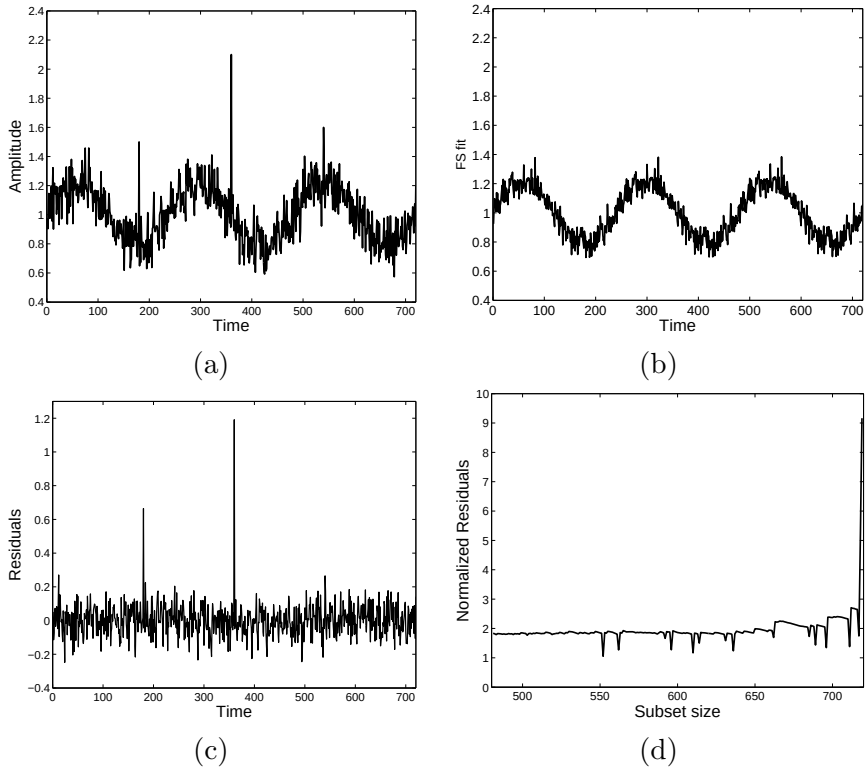
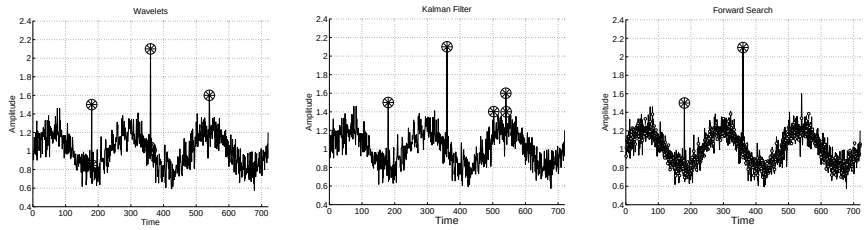


Figure 3: Illustration of Forward Search approach: (a) Data: time series representing the evolution of a quantity over time; (b) Model fitted by the Forward Search; (c) Residuals (difference between data and FS fit); (d) Monitoring of the Forward Search: evolution of the studentized residuals with respect to subset size.

2.5 Conclusion

We presented three methods that are relevant for the automatic extraction of features within trade data, analyzed as time series. All three allow to identify instants at which there is a non-regular behavior and give additional information on the underlying pattern: strength and regularity with WL, prediction and error with KF, and fitted model and evidenced outliers with FS. In the following section we present computational issues, qualitative aspects and a quantitative comparison.



(a) (b) (c)

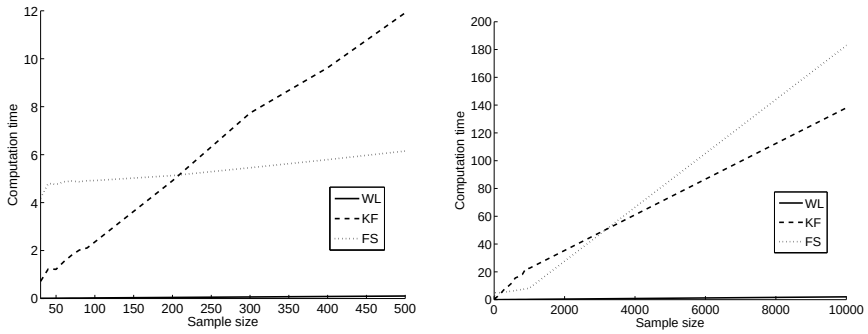
Figure 4: Signals of abnormal behavior detected on data consisting of a seasonal trend, a Gaussian noise and three outliers, using: (a) WL (parameter $T_{WL} = 5$), (b) KF (parameter $T_{KF} = 3$), (c) FS (parameter $T_{FS} = 0.01$).

3 Comparison of the different approaches

To better highlight the differences between the present three methods, we present in Table 1 an overview of their main characteristics. Since the presented methods WL, KF and FS are based on different frameworks, methodological issues appear when one wants to compare them. Therefore it is essential to use a rigorous methodology that does not favor one method *a priori*. We first discuss their computational performance, from the theoretical and practical points of view. Then we compare them qualitatively by illustrating results on classical patterns and commenting on the specificity of each method. Finally we put forward procedures to compare them quantitatively, identifying times that are simultaneously detected by one, two or three methods. With a view of assessing performance on large datasets – on both simulated and real data – these procedures are carried out on each time series, and then we define indicators bearing on the whole dataset.
Important note: all three methods allow to tune the number of signals detected, either for one time series or for the whole dataset. In our experiments we tune them to make sure that for the whole dataset, the number of signals is the same for WL, KF and FS. This allows to obtain a similar number of signals by the three methods for each time series, as represented on Figure 4.

Table 1: Overview of used methods: Wavelets, Kalman Filter and Forward Search. For each method we mention its framework, the algorithm we use, its complexity and practical speed.

Method	Wavelets	Kalman Filter	Forward Search
Framework	Time-frequency analysis	Prediction and optimization	Statistical regression and optimization
Algorithm detecting	singularities	sharp changes	outliers
Additional information	Regularity estimation	Normalized error	Measure of outlyingness (pvalue)
Complexity			
n: samples per POD	$O(n \log n)$	$O(n)$	Greater than O(n)
N_{POD}: nb. of POD	$C_{WL} \times N_{POD}$	$C_{KF} \times N_{POD}$	$C_{FS} \times N_{POD}$
Practical speed	Very fast	Moderate	Moderate



(a) (b)

Figure 5: Evolution of the computational time (in seconds) with respect to sample size n. For the three methods (WL, KF and FS) we perform 100 simulations of Gaussian noise and represent average computational times corresponding to sample sizes: (a) up to 500; (b) up to 10,000. Remark: for WL it is very close to the abscissae axis.

3.1 Computational aspects

In the context of trade data one is often confronted with huge datasets. Hence efficient algorithms are highly desirable to process them in a reasonable time. We presented in section 2 known efficient algorithms for WL, KF and FS – corresponding to the state-of-the-art. Let us now analyze their theoretical complexity and practical speed.

Theoretical complexity. Denoting N_{POD} the number of time series, the complexity of the three methods can be written as $C_{method} \times N_{POD}$. Besides, since data is processed independently for each time series, it is possible to speed up these computations using parallel computing techniques. Now, considering one time series, let us study the complexity and computational times of the different methods. We denote n the number of observations of one time series. Concerning wavelets, we use here the Continuous Wavelet Transform (CWT) It can be efficiently computed by a spectral method (using fast Fourier transforms) of $O(n \log n)$ complexity. Once the CWT has been computed, all operations (such as regularity estimation) can be performed in $O(n)$. So the complexity of our wavelet approach is $O(n \log n)$. Concerning the Kalman Filter, it takes $O(n)$ operations to initialize the state and $O(n)$ to compute prediction and other values at each time, resulting in a complexity $O(n)$. Concerning the Forward Search, its algorithm relies on linear regressions (such as least median squares) and optimization procedures. We recall that the complexity of a simple linear regression is $O(pn^2)$ (p: number of parameters). Given that the computation of the initial subset – ideally free of outliers – entails an enumeration of many possible subsets, and even all for a small data size. In the worst case it can result in a very high complexity (exponential), and in any case it remains greater than $O(n)$. This could be improved using heuristic approaches to provide faster algorithms of FS.

Practical comparison. To evaluate the relative speed of the WL, KF and FS methods, we perform computations for different values of sample size n (from 10 to 10,000). For each value of n, we consider 100 datasets made of Gaussian noise. We represent on Figure 5 average computational times. Besides we compute Taylor expansions of the computational time as a function of the sample size n: denoting ct_{WL}, ct_{KF} and ct_{FS} the computational times (in seconds) of WL, KF and FS, we obtain the following approximations for $n \leq 500$

$$\begin{array}{rcl} ct_{WL} & \approx & (32 + 0.2 \cdot n)/1000 \\ ct_{KF} & \approx & (82 + 24 \cdot n)/1000 \\ ct_{FS} & \approx & (4624 + 3 \cdot n)/1000 \end{array} \tag{17}$$

and also for $n \geq 1000$

$$\begin{array}{rcl} ct_{WL} & \approx & 0.2 \cdot n/1000 \\ ct_{KF} & \approx & 13 \cdot n/1000 \\ ct_{FS} & \approx & 18 \cdot n/1000 \end{array} \tag{18}$$

Since these computational time depend on the machine used, we underline that the important aspect lies in the relative practical speed of the methods. Globally, the performance of KF and FS are of the same order of magnitude, KF being faster than FS for $200 < n < 3000$. For smaller values of n, we note that FS is more influenced than WL and KF by the constant term seen in equation (17). This can be explained by the importance of the initialization step in the Forward Search algorithm (initial subset free of outliers). Overall the WL method is dramatically faster than KF and FS. In particular for larger values of n – see equation (18) – WL turn out as 65 times faster than KF and 90 faster than FS. This huge difference is explained by the direct computations in WL, compared to costly optimizations in KF and FS.

Conclusion. We assessed the computational efficiency of the methods WL, KF and FS, considering associated known efficient algorithms. From the theoretical point of view, all three methods present a complexity suitable to process massive datasets. From the practical point of view, we note a moderate speed for KF and FS whereas the WL algorithm appears as clearly faster.

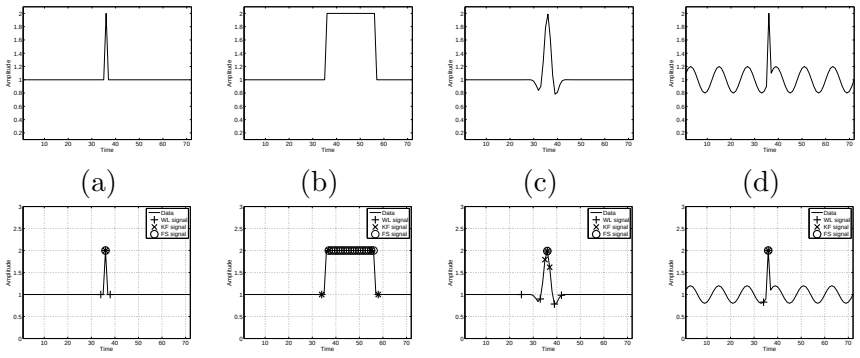


Figure 6: Classical patterns corresponding to abnormal behavior: (a) Sharp spike; (b) Sharp level-shift; (c) Singular waveform localized in time; (d) Spike on a smooth seasonal evolution. (a',b',c',d') Detected signals of abnormal behavior using WL, KF and FS.

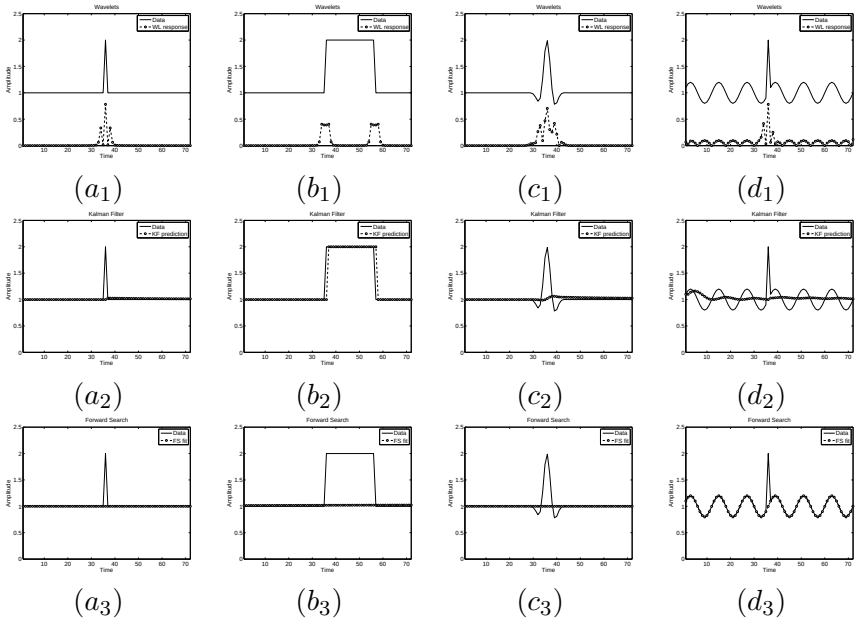


Figure 7: For each pattern $i = 1..4$ seen in Fig.6, spike, level-shift, localized waveform, seasonal trend with a spike: (a_i) WL response, (b_i) KF prediction, (c_i) FS fit.

3.2 Qualitative comparison

The presented methods are designed to detect specific patterns. Since each method has its own philosophy and underlying assumptions on data, we expect them to perform differently depending on the abnormal pattern present in the data. We represent some classical patterns on Figure 6. Let us precise which patterns are likely to be detected by WL, KF and FS. The wavelet approach uses a multiscale transform of the data to detect singular patterns. This makes it well adapted to high-frequency waveforms, spikes and level-shifts. The Kalman Filter allows to predict the evolution of the data according to previous observations, and can be used to detect sudden changes (spikes, level-shifts). The Forward Search allows to detect outliers in various contexts, in particular times series having a seasonal trend. It is well adapted to cases when data behaves as a clear seasonal evolution except for a few outliers. In addition to the detection of patterns at certain times, these methods provide additional information on these patterns. Concerning WL, it is possible to compute response and values of regularity. These quantify the strength and the sharpness of a pattern, and allow to classify it as spike, level-shift or more regular. Concerning KF, this method provides normalized error, which quantifies to which degree an observation is far from the KF prediction. Concerning FS, once the outliers have been identified, an ad-hoc formula allows to compute residuals. In the context of large datasets to be processed, an important aspect is to control the number of detected patterns. This can be done through thresholding on an appropriate parameter, which quantifies the strength of a pattern: measure of strength for WL, normalized residuals for KF and FS. Although these measures differ from one method to another (and cannot be compared directly), they allow to tune the number of detected patterns.

Let us now compare results obtained on simple examples by these approaches (WL, KF and FS). We represent on Figure 6 classical patterns of abnormal behavior within time series, often encountered in practice when dealing with trade data. Applying the methods WL, KF and FS (using adequate thresholding), we detect signals corresponding to these patterns. The spike pattern is effectively detected by all three methods, and WL also detects singularities before and after the spike. The level-shift pattern is detected by WL and KF but not by FS (no clear outlier). Concerning the waveform pattern, WL detects several singularities, KF one sharp change, and FS one clear outlier. Finally the pattern consisting on a spike over the seasonal trend is well detected by all three methods. Additionally we represent on Figure 7 the WL response, the KF prediction and the FS fitted model corresponding to these four classical patterns. This allows to understand better how these methods detect signals of abnormal behavior. First, WL focuses on pointwise singularities, sometimes leading to several signals for one pattern. Second, KF focuses on sharp changes and performs best when there is one sudden change over a smooth trend (spike, level-shift). Third, FS focuses on outliers using linear or seasonal models, yet it does not perform well for level-shifts and non-stationary time series. In conclusion, all three methods effectively identify these classical patterns, using different approaches for signal detection.

3.3 Quantitative comparison methodology

We present here methodologies that allow to evaluate the performance of the used methods. We address two cases:

- Case of simulated data: we apply the proposed methods to times series containing patterns seen in section 3.2. Since signals of abnormal behavior are known, we can compare them with the signals detected by each method WL, KF and FS. This allows to evaluate the performance of each method, comparing practical results to theoretical ones.

- Case of real data: we apply the proposed methods to times series relative to EU trade data, as described in section 1.2. In this context where signals of abnormal behavior are *a priori* unknown, we rely on inter-method comparison. This allows to evidence the common aspects and the differences between WL, KF and FS.

Case of simulated data. We carry out the following procedure

1. For every time series TS_i, $i = 1..N$ (N: number of POD/time series), each method M (here M =Wavelets, Kalman Filter or Forward Search) detects a set of time instants

$$S_i(M) = \{t_i^{i,m}\} \tag{19}$$

2. For every time series, compute the matching score defined as

$$MS_i(M) = \frac{|S_i(M) \cap S_i^0|}{\max(1, |S_i(M)|, |S_i^0|)} \in [0, 1] \tag{20}$$

   where S_i^0 is the set of time instants corresponding to a known pattern.

Now, considering all time series, we compute the average matching score $MS(M) = \frac{1}{N} \sum_i MS_i(M)$. This measures to which degree one method is adapted to a given pattern. Additionally we compute the rate of effective detection, defined as the percentage of time series for which the pattern was detected (case when $S_i(M) \cap S_i^0 \neq \emptyset$). This measures the ability of one method to detect effectively signals of abnormal behavior.
Case of real data. We carry out the following procedure

1. Perform step 1 used for simulated data to obtain sets $S_i(M)$ for each time series and each method

2. For every time series, compute the matching score defined as

$$MS_i(M, M') = \frac{|S_i(M) \cap S_i(M')|}{\max(1, |S_i(M)|, |S_i(M')|)} \in [0, 1] \tag{21}$$

   which measures common signals of abnormal behavior between two methods M and M'.

We use this procedure to compare on the whole dataset the average matching scores between two methods (WL and KF, WL and FS, KF and FS).

Comments. Let us present extreme cases, to better interpret the meaning of the matching score. First, let us consider a method M that identifies all times as abnormal for any time series. In such case equation (20) becomes

$$MS_i(M) = \frac{|S_i(M) \cap S_i^0|}{\max(1, |S_i(M)|, |S_i^0|)} = \frac{|S_i^0|}{N} \tag{22}$$

Besides, considering such methods M and M', equation (21) becomes

$$MS_i(M, M') = \frac{|S_i(M) \cap S_i(M')|}{\max(1, |S_i(M)|, |S_i(M')|)} = \frac{N}{N} = 1 \tag{23}$$

Hence, in such case the score $MS_i(M)$ is low (low performance) while the score $MS_i(M)$ is high (100% common part). Now, considering again such method M and a perfect method M' giving $S_i(M') = S_i^0$, equation (21) becomes

$$MS_i(M, M') = \frac{|S_i(M) \cap S_i(M')|}{\max(1, |S_i(M)|, |S_i(M')|)} = \frac{|S_i^0|}{N} \tag{24}$$

which results in a low score. This illustrates the relevance of these matching scores to evaluate the performance of one method, or to assess the common part between two methods.

Important note: for each method, one can control the number of extracted signals using an adequate parameter. Although the matching scores previously defined allow to compare very different methods, it can be useful to tune the parameters so that the number of detected signals are roughly similar. Besides, for each method used here one can rank the detected signals by importance thanks to an adequate measure (response for WL, error for KF, outlyingness for FS). This can be used to obtain exactly the same number of signals for each method.

4 Results

We present here results obtained on simulated and real data. Considering one dataset made of time series, we first detect signals of abnormal behavior using all three methods (WL, KF and FS). Then we apply on these signals the methodologies presented in section 3.3. Results allow to draw conclusions on the relevance of each method to detect certain patterns.

4.1 Application on simulated data

Experiments. We apply the methods WL, KF and FS on the sum of one pattern and a Gaussian white noise. The pattern is one of those represented on Figure 6: spike, level-shift, localized waveform, or spike over a trend. We add Gaussian noise to perturbate these clear signals of abnormal behavior, using the following signal-to-noise ratios: 20dB, 15dB, 10dB and 5dB – we recall the signal-to-noise ratio is defined as $SNR = 20 \log_{10} \frac{\text{Noise amplitude}}{\text{Signal amplitude}}$ and that a lower SNR means a higher noise level. For instance a 20dB SNR means the signal is 10 times stronger than the noise. Note this differs from other contamination methods used in robust statistics, such as adding outliers randomly. This would be inappropriate here because such outliers would be detected as abnormal patterns. Every method is applied on 100 simulations of each pattern and each noise level. Denoting S_0 the known location of the pattern, we compute the matching score $S(M, S_0) \in [0, 1]$ for each method and each pattern (average value over 100 simulations). Moreover we compute the rate of effective detection of the pattern, defined as the percentage of simulations for which the known pattern was effectively detected (average value over 100 simulations). A method is all the more efficient than these two values are high: ideally the matching score should be close to 1, and the effective detection close to 100%. Results are summarized in Table 2.

Table 2: **Results on simulated data**. Matching score and effective detection rate obtained by applying our methodology to data made of noisy classical patterns (the level of noise is quantified by the SNR, high SNR corresponding to low noise).

	Matching score			Effective detection		
SNR	WL	KF	FS	WL	KF	FS
20dB	0.66	0.50	1.00	100%	100%	100%
15dB	0.52	0.31	0.90	99%	100%	90%
10dB	0.43	0.26	0.19	76%	100%	19%
5dB	0.24	0.19	0.02	41%	81%	2%

(a) Spike

	Matching score			Effective detection		
	WL	KF	FS	WL	KF	FS
20dB	0.75	0.66	1.00	99%	100%	100%
15dB	0.73	0.51	0.90	94%	100%	90%
10dB	0.52	0.43	0.19	68%	100%	19%
5dB	0.37	0.38	0.02	48%	95%	2%

(b) Level-shift

	Matching score			Effective detection		
	WL	KF	FS	WL	KF	FS
20dB	0.51	0.67	1.00	100%	100%	100%
15dB	0.55	0.42	0.85	100%	100%	85%
10dB	0.48	0.29	0.12	81%	100%	13%
5dB	0.24	0.24	0.02	40%	97%	2%

(c) Localized waveform

	Matching score			Effective detection		
	WL	KF	FS	WL	KF	FS
20dB	0.52	0.35	1.00	100%	100%	100%
15dB	0.53	0.31	0.82	100%	100%	82%
10dB	0.42	0.27	0.12	76%	99%	12%
5dB	0.19	0.22	0.02	36%	87%	2%

(d) Spike over a seasonal trend

Results. First let us analyze the results for each pattern. Concerning the spike pattern – see Tab 2(a))– FS performs very well for low level of noise, and poorly for higher noise levels. WL and KF perform quite well overall, and appear as more robust to higher noise levels (compared to KF). We observe similar results for the level-shift pattern – see Tab 2(b). For low noise level, FS performs very well while WL and KF perform well. For higher noise level, WL and KF perform better and appear as more robust than FS. In particular we note that for both spike and level-shift patterns, WL possess better accuracy (matching score) compared to KF and FS. Concerning the localized waveform pattern – see Tab 2(c) – results for WL and KF are good overall, while FS performance decreases sharply as the level of noise increases. Finally we note that WL obtain a good matching score overall (which underlines its robustness) while KF maintains a high effective detection rate. Concerning the spike over a seasonal trend pattern, – see Tab 2(d) – we note that for low levels of noise, FS performs significantly better than WL and FS (considering matching scores). However the performance of FS drastically drops for higher levels of noise. In contrast, WL and KF present a better robustness to noise, especially KF in terms of effective detection rate.

Now let us analyze the results on Table 2 focusing on the impact of the noise level. In the context of low level of noise (20dB, 15dB), FS turns out as very efficient, with a matching score close to 1 and an effective detection rate close to 100%. WL and KF perform well overall: matching score good for WL and average for KF, while both methods attain a high effective detection rate. In the context of a high levels of noise (10dB, 5dB), the performance of FS falls dramatically while WL and KF present a better robustness. More precisely WL obtain slightly better matching score than KF whereas KF attains a better effective detection rate (especially for the highest level of noise). Finally we underline that in the context of high level of noise, the results obtained by WL and KF can be considered as good: even visually the patterns are difficult to distinguish from the noise.

Conclusion. We presented here results on simulated data for which the pattern are known *a priori*, using different levels of noise. These results show that FS performs very well in the context of low noise level, and poorly for high level of noise. WL perform well overall, having a good robustness to noise. Finally KF perform well overall, having a very good robustness to noise.

Table 3: **Results on real data**. We apply the methods WL, KF and FS on a dataset reporting EU trade data, consisting of 26,233 times series. We first compare the detected signals of the three methods for each time series (here $K = 5000$) and then synthesize the results for the whole dataset: (a) Matching score, as defined in our methodology for real data; (b) Percentage of perfect matches, defined as POD for which the signals detected by two methods are identical.

Matching score

	WL	KF	FS
WL	1	0.25	0.21
KF	-	1	0.30
FS	-	-	1

(a)

Percentage of perfect matches

	WL	KF	FS
WL	100%	19%	12%
KF	-	100%	23%
FS	-	-	100%

(b)

4.2 Application on real data

Experiments. Here we consider real data extracted from the huge external trade database mentioned in the introduction (COMEXT, Eurostat). We focus in particular on imports of products entering the EU, spanning the period 2008-2011, trade being reported monthly, and bearing on all 99 chapters of the Integrated Tariff of the European Communities (TARIC). We underline this makes up a comprehensive dataset on trade between EU and third countries, covering a very large set of traded products. In terms of dimensions, this dataset contains 16 million records. In summary data consist of a large number of time series with 36 observations, each reporting the volume of a product over time. We explained in section 1 the relevance of the detection of abnormal behavior within such data. We apply the methods WL, KF and FS on each time series. This is carried out independently for each method, using adequate thresholds, we ensure that the number of detected signals are equal for each method. We denote $K \in \mathbb{N}$ this number of most important signals to be detected.

$$K = N_{signals}(WL) = N_{signals}(KF) = N_{signals}(FS) \tag{25}$$

These signals of abnormal behavior obtained on real data make up a basis to compare the three methods. Applying the methodology presented in section 3.3, we compute matching scores between two methods (WL vs. KF, WL vs. FS, KF vs. FS). We report in Table 3(a) average values on the whole dataset. This allows to evaluate to which extent signals of

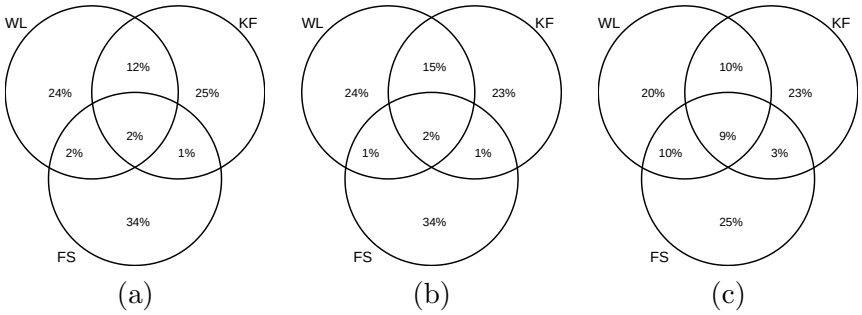


Figure 8: **Results on real data**. Common parts and differences between signals detected by WL, KF and FS. We apply the three methods WL, KF and FS on a dataset reporting EU trade data. For each method we extract a subset containing the K most important signals: (a) K=100; (b) K=1000; (c) K=5000. Considering the whole set made of these three subsets, we identify the signals of different methods corresponding exactly the same time of the same time series. This allows to classify signals into the following categories: those detected by only one method (WL for instance) only two methods (WL and KF for instance) and by all three methods (WL, KF and FS).

one method correspond to those of another method. We point out matching scores should be in any case lower for real data compared to simulated data: denoting T the set of (unknown) ground truth, $MS(M, M') \leq \min(MS(M, T), MS(M', T))$. Furthermore we identify the time series for which two methods evidence exactly the same time instants (as signals of abnormal behavior). Then we define the percentage of perfect matches as the proportion of such time series among the time series for which at least one signal was detected by any of the two methods. This actually identifies the percentage of the time series for which the matching score equals one. We report in Table 3(b) the percentage of time series for which there is such perfect match. Finally we identify the common parts and the differences between the signals detected by the three methods. More precisely, for each time series we consider the time instants identified as abnormal by any of the three methods. Then, for each time series and each signal we check if it corresponds to: only WL, only KF, only FS, only two of them, or all three of them. Considering now the whole dataset, we apply these steps selecting the K most important signals for each method. Results provide a global picture of the signals of abnormal behavior, see Figure 8.

Results. Let us analyze the results of Table 3. We recall the matching score measures the similarity between signals detected by one method and those detected by another, and that reported values are average over the whole time series for which there was (at least) one signal detected. In Table 3(a), we note relatively low values of matching score overall, between 0.2 and 0.3, the lowest corresponding to WL-FS and the highest to KF-FS. So on real data the three methods produce heterogeneous results. This can be explained by the fact they are based on different frameworks, as explained in section 3.2. Empirically, when we look at the corresponding time series, we note that this is often due to a lack of precision (like several singularities detected by WL when there is only one signal), the abnormal pattern being still effectively detected. Concerning perfect matches, we observe values between 12% (WL-FS) and 23% (KF-FS). Given the strict criterion of perfect match, these relatively moderate values indicate actually a certain correspondence between the results: there are patterns for which all three methods give very similar results, as we previously noted on simulated data.

Now let us analyze the results of Figure 8. Choosing a certain value of K (in our experiments $K = 100, 1000, 5000$) we obtain a set of most important signals according to all three methods (less than $3 \cdot K$ given the common parts). We can then identify common parts and differences within this set. Globally we first note that results for $K = 100$ and $K = 1000$ are very similar, and that there is a greater part in common for $K = 5000$ (compared to $K = 1000$ and $K = 100$). Thus when selecting a larger number of signals, one gets signals which are more likely to be detected by two or three methods rather than one. Second, focusing on the most important signals ($K = 100$), as reported on Figure 8(a), we observe that 2% of the signals correspond to all three methods, 1% to KF and FS, 2% to WL and FS, 12% to WL and KF. Thus we note a larger part in common between WL and FS. This can be explained by the link between singularities (WL approach) and sharp changes (KF approach). Finally we note that there is a wide part of signals that are detected by only one method, even when selecting the most important signals: for $K = 100$, this proportion of signals attains 83% – 24% for WL, 25% for KF, 34% for FS. (other experiments with $K = 10$ lead to a proportion of 95%). We also point out that restricting the number of selected signals results in lower common parts – each method focusing on its reference pattern. Considering time series, we note that roughly 20% of them correspond to perfect matches – corresponding generally to simple patterns with low noise.

In summary, we note a small proportion of signals is simultaneously detected by all three methods (between 2% and 9%), a moderate part is detected by at least two (between 15% and 23%), and the largest part is detected only by one (between 68% and 83%). This can be explained by the different frameworks of the three methods, and the variety of patterns encountered in real data.

Qualitative analysis. In addition to the presented quantitative results, let us present qualitative aspects. We underline no ground truth is known for such huge datasets, and that expert evaluation is often subject to confidentiality. We perform an empirical study on time series, assessing visually the computed results by a group of non-specialists. More precisely we consider two cases: when at least one signal was detected in a time series, either by each method (first case) or by any of the methods (second case). In both cases we consider 100 time series (randomly chosen), and we retain the consensus opinion of the group: good (pattern precisely detected), satisfying (pattern approximately detected, several signals when there is only one pattern), poor (main pattern missed, or signal detection when there is no clear pattern). We represent in Table 4 a synthesis of these results. First no significant difference is seen between Table 4(a) and (b), which indicates a certain stability of the methods. Second, the obtained results are very good for WL and KF – and also similar. Given the ambiguity of certain time series, this emphasizes the relevance of these methods. Third, we note average results for FS. This actually comes from the fact FS performs poorly when no seasonal trend can be successfully fitted. However it gives good results when the time series presents a clear seasonal trend. Let us detail comments on each method.

Wavelets: very good overall, in particular concerning level-shifts and spikes. Nevertheless we note a tendency to detect several singularities when only one pattern is present.

Kalman Filter: very good overall, especially for upward spikes. We note a tendency to erroneously detect signals when time series are very oscillating. Besides some level-shifts and downward spikes are not always detected.

Forward Search: good when a seasonal fit is successful. However, when only a linear fit is performed, it performs poorly (either zero or too many outliers).

Table 4: **Results of qualitative evaluation**. Proportion of time series for which the signal detection by WL, KF and FS was evaluated as good, satisfying or poor (according a group consensus). The considered time series obey the following rules: (a) at least one signal was detected by WL, KF and FS; (b) at least one signal was detected by WL, KF or FS.

	WL	KF	FS
Good	75%	72%	51%
Satisfying	21%	21%	28%
Poor	4%	7%	21%

(a)

	WL	KF	FS
Good	74%	70%	52%
Satisfying	20%	21%	30%
Poor	6%	9%	18%

(b)

Conclusions. We presented here results obtained on real data containing a variety of patterns. We performed a quantitative evaluation of the signals detected, which allowed to compare the methods. Further analysis was provided on qualitative aspects. This provides better understanding of the results obtained by these methods on real data, and in which cases such or such method turns our as efficient. The obtained results suggest the joint use of these three methods can be relevant for the applications. For instance, we can classify signals of abnormal behavior into three classes using the following criterion: provided a time instant was detected (respectively) by one, two or three of the presented methods, it can be identified as a light, medium or strong signal. Considering the studied real data (for $K = 100$) we obtain respectively 83% of light, 15% of medium, and 2% of strong signals of abnormal behavior. Such an approach would then allow the classification of the time series, based on the number of signals detected by each method.

4.3 Recommendations – Method choice

Here we detail practical recommendations on the use of such or such method, according to the results obtained on simulated and real data.

Identification of sharp changes. When the task consists in identifying sharp changes (spikes, clear level-shifts) in time series, WL and KF proved their efficiency (FS performed poorly except for simple patterns and low noise level). So in this case we recommend first KF and second WL.

Distinguishing spikes, level-shifts and more regular patterns. In this case, the WL presents an added value: the estimated value of regularity allows to distinguish how regular a pattern is (what KF and FS cannot do). So when the user wishes to separate signals depending on their type (only level-shifts for instance), we recommend WL.
Identification of extra-seasonal breaks. In the context of time series presenting a clear seasonal trend, some of them present an extra-seasonal break: low or high value at one month, compared to quantity of the preceding years at the same month. In such cases FS shows its very relevance, only identifying these extra-seasonal breaks as outliers. In contrast, KF detects high values but not low values, while WL tend to detect all variations. So for practical applications in which products are naturally subject to such seasonal trends (agricultural chapters for instance), we recommend the FS approach, and KF to a certain extent.
In-depth and thorough analysis. We previously pointed out the possible joint use of WL, KF and FS. This is valid when the user wants to carry out an in-depth analysis of a limited dataset. The added value of this joint use is to focus on signals that were detected: by all three methods (strong signals); simultaneously by WL and KF whilst not by FS (spikes); and only by FS (extra-seasonal breaks).
Processing of massive datasets. While in theory all three methods have a similar computational complexity, we note significant differences in practice: while KF and FS lead to satisfying computational times, WL turns out as extremely fast. Hence for huge datasets (several million of records, data size greater than one gigabyte) we recommend the WL approach.

5 Conclusions and perspectives

In this paper we addressed a topic of interest for applications such as antifraud and fight against money-laundering. We studied methods allowing the detection of abnormal behavior within time series: Wavelets, Kalman Filter and Forward Search. These are based on different frameworks (applied mathematics, signal processing and statistics), and each has its specific features. We highlighted their differences and carried out a qualitative comparison. An original contribution of this work is a general methodology

allowing to effectively compare these methods, on both simulated and real datasets. Concerning simulated data, results show that all three methods can precisely detect classical patterns (the extracted signals being equal in simple cases). The wavelet approach performs well overall, with a good robustness to noise. The Kalman filter approach performs well, with a very good robustness to noise. The Forward Search approach performs well for low level of noise, but poorly for higher levels of noise. Concerning real data reporting EU trade, results show that the signals detected by the three methods present some common part and also significant differences. These greater differences on real data (compared to simulated data) are explained by the greater complexity of real data (compared to simulated one). Finally we noted the great computational efficiency of the wavelet approach, in absolute terms and also compared to Kalman Filter and Forward Search.

In terms of applications, the joint use of these three approaches can turn out as relevant for the monitoring of EU trade data. This is emphasized by the additional information these methods give on the underlying pattern: wavelet response and local regularity estimation, Kalman Filter prediction and error, fitted model and residual obtained by applying the Forward Search. Hence a signal of abnormal behavior simultaneously detected by three methods could be considered as relevant for further analysis by antifraud analysts or specialists of the fight against money-laundering. More generally data monitoring systems can benefit from such integrated approaches, especially when the used methods present some complementarity. Finally when these methods provide further interpretable information (such as the local regularity in the wavelet approach), one better understands why and how such abnormal pattern was detected. This motivates the joint use of these methods in the applications.

Finally let us mention some perspectives on the presented methods. First, enhancing the multiscale aspect in the wavelets approach would allow to detect patterns more robustly. Second, the Kalman Filter could be improved using a robust formulation (reducing the effect of outliers on the prediction). Third, the Forward Search could be improved in general by efficient algorithms (heuristic approaches without full search), and particularly for times series by allowing more flexibility concerning the model (adaptive seasonality).

Acknowledgment

This work was supported by the European Commission, under the institutional program of the action SITAFS (Statistics and Information Technologies for Antifraud and Security). The author would like to thank for their useful comments and suggestions L. Orfei and D. Perrotta (European Commission, Joint Research Centre, Ispra, ITALY).

References

Abraham B, Chuang A (1989) Outlier detection and time series modeling. Technometrics 31(2):241–248

Andersson P (1997) Characterization of pointwise hlder regularity. Applied and Computational Harmonic Analysis 4(4):429–443

Atkinson A, Riani M (2000) Robust Diagnostic Regression Analysis. Computational Statistics

Atkinson A, Cerioli A, Riani M (2004) Exploring Multivariate Data With the Forward Search Springer Series in Statistics. Springer

Barnett V, Lewis T (1994) Outliers in statistical data, 3rd ed. Wiley, New York

Basu S, Meckersheimer M (2007) Automatic outlier detection for time series: an application to sensor data. Knowledge and Information Systems 11(2):137–154

Benassi A, Cohen S, Istas J, Jaffard S (1998) Identification of filtered white noises. Stochastic Process Appl 75(1):31–49

Beylkin G, Coifman R, Rokhlin V (1991) Fast wavelet transforms and numerical algorithms. Communications on Pure and Applied Mathematics XLIV:141–183

Bolton R, Hand D (2002) Statistical fraud detection: A review. Statistical Science, 17(3)

Caudell T, Newman D (1993) An adaptive resonance architecture to dene normality and detect novelties in time series and databases. Proceedings of IEEE Neural Networks

Chandola V, Banerjee A, V K (2009) Anomaly detection: a survey. ACM Computing Surveys, NY, USA 41(3)

COMEXT (Eurostat) External trade database for the european union. Eurostat (Statistical Office of the European Union), http://epp.eurostat.ec.europa.eu/newxtweb

Damerval C, Meignen S (2009) Study of a robust feature: The pointwise lipschitz regularity. International Journal of Computer Vision 88(3):363–381

Donoho D, Johnstone I (1994) Ideal spatial adaptation by wavelet shrinkage. Biometrika 81:425–455

Durbin J, Koopman S (2001) Time series analysis by State Space Methods. Oxford university press

Fawcett T (1997) Ai approaches to fraud detection and risk management. Association for the Advancement of Artificial Intelligence workshop

Fogelman-Soulie F, Perrotta D, Piskorski J, Steinberger R (eds) (2008) Mining Massive Data Sets for Security. NATO Science for Peace and Security, Volume 19, Series - D: Information and Communication Security, Advances in Data Mining, Search, Social Networks and Text Mining, and their Applications to Security

Fox A (1972) Outliers in time series. Journal of the Royal Statistical Society Series B 34:350363

FSDA (2011) Matlab toolbox on forward search and data analysis. http://www.riani.it/MATLAB.htm

Geurts P (2001) Pattern extraction for time series classification. In: Principles of Data Mining and Knowledge Discovery, Lecture Notes in Computer Science, Springer

Hoaglin D, Mosteller F, Tukey J (1983) Understanding Robust and Exploratory Data Analysis, John Wiley and Sons, pp 404–414

Jaffard S, Meyer Y (1996) Wavelet methods for pointwise regularity and local oscillations of functions. American Mathematical Society

Kalman R (1960) A new approach to linear filtering and prediction problems. Journal of Basic Engineering, Transactions ASMA, Series D 82:33–45

Keogh E, Lonardi S, ChiChiu B (eds) (2002) Finding surprising patterns in a time series database in linear time and space

Ma J, Perkins S (2003) Time series novelty detection using one-class support vector machines. In: International Conference on Neural Networks, vol 3

Mallat S (1989) A theory for multiresolution signal decomposition: The wavelet representation. IEEE Transactions on Pattern Analysis and Machine Intelligence 11:674–693

Mallat S, Hwang W (1992) Singularity detection and processing with wavelets. IEEE Transactions on Information Theory 38(2):617–643

Peng J, Aston J (2007) State space models, matlab toolbox on kalman filtering. http://sourceforge.net/projects/ssmodels

Peng J, Aston J (2010) The state space models toolbox for matlab. Journal of Statistical Software, 41(6):1–26, http://www.jstatsoft.org/v41/i06/paper

Riani M (2004) Extensions of the forward search to time series. Studies in Nonlinear Dynamics and Econometrics 8(2)

Rousseeuw P, Leroy A (1987) Robust regression and outlier detection. Wiley, New York

Salvador S, Chan P (2005) Learning states and rules for detecting anomalies in time series. Applied Intelligence 23(3):241–255

Soule A, Salamatian K, Taft N (eds) (2005) Combining ltering and statistical methods for anomaly detection, ACM

Strang G, Nguyen T (1996) Wavelets and filter banks. Wellesley-Cambridge Press, USA

Wavelab (1992) Matlab toolbox on wavelet analysis. `http://stat.stanford.edu/~wavelab`

European Commission
EUR 25491 EN Joint Research Centre – Institute for the Protection and Security of the Citizen

Title: **Detection of abnormal behavior in trade data using Wavelets, Kalman Filter and Forward Search**

Authors: Christophe Damerval

Luxembourg: Publications Office of the European Union

2012 – 40 pp. – 21.0 x 29.7 cm

EUR – Scientific and Technical Research series – ISSN 1831-9424 (online), ISSN 1018-5593 (print)

ISBN 978-92-79-26265-4

doi:10.2788/46203

Abstract

In this paper we address the issue of the automatic detection of abnormal behavior in time series extracted from international trade data. We motivate, review and use three specific methods, based on solid frameworks: Wavelets, Kalman Filter and Forward Search. These methods have been successfully applied to an important EU policy issue: the analysis of trade data for antifraud and antimoney-laundering, fields in which specialists are often confronted with massive datasets. Our contribution consists in an in-depth study of these approaches to assess their performance, qualitatively and quantitatively. On the one hand, we present these three approaches, underline their specific aspects and detail the used algorithms. On the other hand, we put forward a rigorous assessment methodology. We use this methodology to evaluate each method and also to compare them, on simulated time series and also on real datasets. Results show each method has its specific advantages. Their joint use could be of a high operational impact for our applications, to deal with the variety of patterns occurring in trade data.

LB-NA-25491-EN-N

As the Commission's in-house science service, the Joint Research Centre's mission is to provide EU policies with independent, evidence-based scientific and technical support throughout the whole policy cycle.

Working in close cooperation with policy Directorates-General, the JRC addresses key societal challenges while stimulating innovation through developing new standards, methods and tools, and sharing and transferring its know-how to the Member States and international community.

Key policy areas include: environment and climate change; energy and transport; agriculture and food security; health and consumer protection; information society and digital agenda; safety and security including nuclear; all supported through a cross-cutting and multi-disciplinary approach.

ISBN 978-92-79-26265-4

9 789279 262654

Publications Office